RAID ASSOCIATIVE TOOL REQUIREMENTS SPECIFICATION

MANSUR DARLINGTON, ALEX BALL, TOM HOWARD, STEVE CULLEY AND CHRIS MCMAHON

erim6rep101111mjd10.pdf

ISSUE DATE: 23 FEBRUARY 2011

Catalogue Entry

Title	RAID Associative Tool Requirements Specification
Creator	Mansur Darlington, Alex Ball, Tom Howard (authors)
Contributor	Steve Culley, Chris McMahon
Subject	information management; data management; research data; software requirements
Description	Association is a data development activity that greatly improves the ability of secondary users to contextualize and understand research data. It is of such importance that the ERIM Project proposes a software tool to assist in the construction of associative data records, specifically using the Research Activity Information Development (RAID) modelling method. The design requirements for such a tool are presented, with the focus on the context of its use, the user tasks the tool must support, and the functions that the tool must consequently perform.
Publisher	University of Bath
Date	11th November 2010 (creation)
Version	1.0
Type	Text
Format	Portable Document Format (PDF/A-1b:2005)
Resource Identifier	erim6rep101111mjd10
Language	English
Rights	

Citation Guidelines

Mansur Darlington, Alex Ball, Tom Howard, Steve Culley and Chris McMahon. (2011). *RAID Associative Tool Requirements Specification* (version 1.0). ERIM Project Document erim6rep101111mjd10. Bath, UK: University of Bath.

1. INTRODUCTION

The desire to increase the opportunity for re-using existing research data that has grown in recent years is exemplified by the research and other activities supported by and carried out by such organizations as JISC, DCC and UKOLN, encouraged and supported in turn by the research funding councils.

This document stems from work carried out by the ERIM (Engineering Research Information Management) Project, which is one of the research data management planning projects within the JISC Managing Research Data Programme.

A central part of the ERIM Project has been to characterize the research data encountered within Engineer Research and to find ways of promoting at the research level ways of supporting the three data preparation activities identified by the authors as:

Data Purposing Making research data available and fit for the *current* research activity.

Data Re-purposing Making existing research data available and fit for a *future known* research activity.

Supporting Data Re-use Managing existing research data such that it will be available for a *future unknown* research activity.

Each of these contributes to the desired and desirable activity of data use and re-use.

As reported in Howard, et al. (2010) one of the greatest impediments to re-using data is the lack of support for recording and maintaining context amongst the diverse digital and paper records which constitute the research **Data Case**(s) developed during the course of research activity. It is argued that without providing some sort of integrating contextualization the understanding necessary for data re-use cannot be guaranteed. In order to provide a good basis for interpreting data, it is necessary to understand the rôles and data content of records and the relationship between data and between records. Such context in which data can be understood is particularly necessary after the event, when the observer of the data has little or no personal knowledge of the circumstances surrounding their original collection or generation.

There are a number of ways of recording context between data including the use of embedded metadata (Sefton, et al., 20089; Kelly, 2004) explanatory annotation (Cobos & Schlichter, 2004) or stand-alone documents. Another approach has, however, been suggested as a result of the work in the ERIM Project. It has been established there that there are data development activities which commonly occur at the **Data**, **Data Record** and Data Case levels during the research activity and which result in the development of research data and the records in which they are found and explained. These activities are identified and defined in Howard, et al. (2010). At the same time, the Research Activity Information Development (RAID) modelling method has been developed by the authors which, in principle, will allow these data development activities to be mapped for a given research activity, such that later interpretation of the data and the data records can be made more easily than is currently the case.

Amongst the data development activities identified is that of *association*, defined as being ‘to make explicit the relationship between items of Data, Data Records or Data Cases’. Seven strategies of association – of increasing completeness – between data and

between data records are identified in Howard, et al. (2010). The RAID modelling method may be used as one means of achieving association through the creation of **Context Data Records** or **Associative Data Records** (See Section 5, Glossary for definitions) and thus provide some part of the contextualization necessary for later data interpretation. It can be shown too, that a RAID model represents the class of associational strategies which provides the most complete association. In such a strategy the association between data or data records is reciprocal or bi-directional.

In summary of the above, as motivation for the development of a tool which will create an associational record, the concluding paragraph of Howard, et al. (2010) is quoted, with specific reference to the sorts and diversity of research data found commonly in engineering research:

In considering the different types of development process to which research data is commonly subjected, the authors believe that 'association' is perhaps foundational in supporting good data management for its easier re-use and re-purposing. To support the researcher in the better management of data as it is developed, and to provide a 'map' to aid its later understanding, the authors propose the use of an automated 'association' tool, based on the information that can be captured and represented in a RAID diagram. By capturing and recording the development of research data in individual activities or projects it is proposed that data assets can be made more findable, interpretable, verifiable, repeatable, replicable and useful'.

The purpose of the content which follows is to provide the *design requirements* for a tool which carries out the process of RAID mapping as data and data records are developed during a research activity. This is the first stage in the life-cycle of software development that might result in such a tool, and which can provide the modelled data in a number of ways. The design requirement provided here consists of two parts: the Requirement Definition (see Section 3) which is an informal description of the application and the context in which it will be used; and the Functional Specification (see Section 4) which identifies the system requirements necessary to provide the services to be provided by the application. This requirements document will 'set out what the system should do without specifying how it should do it' (Pressman, 2005)) and will be based upon the structure of specificational documents recommended by Pressman, and by Fancellu (n.d). The requirements identified and the resultant specification presented here are supported by and should be read in conjunction with the use-case analysis recorded in 'RAID Associative Tool Use Cases' (ERIM Project document erim6rep101125mjd10). The name RAIDmap has been adopted for the application that would result from implementation of the specifications which follow.

2. THE RESEARCH ACTIVITY INFORMATION DEVELOPMENT MODELLING METHOD

A case is made above for recording the development and relations between data and between data records during the course of the research activity. One way of achieving this is by means of the Research Activity Information Development (RAID) Modelling approach presented first in Howard, et al (2010). A brief review of this approach is given here before considering the application requirements definition.

During the course of research data and data records are, in general, collected, generated, manipulated and otherwise developed in one way or another. These 'data development processes' can be seen in Table 1 below in relation to the data preparation activities, for which definitions can be found here.

	Data Preparation for		
Data Development	***Purposing***	***Re-purposing***	***Supporting Re-use***
Addition	✓	×	×
Association	✓	✓	✓
Aggregation	✓	×	×
Annotation	✓	✓	✓
Augmentation	✓	✓	×
Collection	✓	×	×
Collation	✓	✓	✓
Deletion	×	✓	×
Derivation	✓	×	×
Duplication	×	×	×
Extraction	✓	✓	×
Generation	✓	×	×
Migration	✓	✓	×
Population	✓	×	×
Refinement	✓	✓	×

Table 1. Data development and the three data preparation activities

By completion of a research project or other delineated research activity, there are commonly many different data records which are related in one way or another to the research effort and which have been created through the data development activities. These can be seen in Figure 1, and are defined in Section 5.

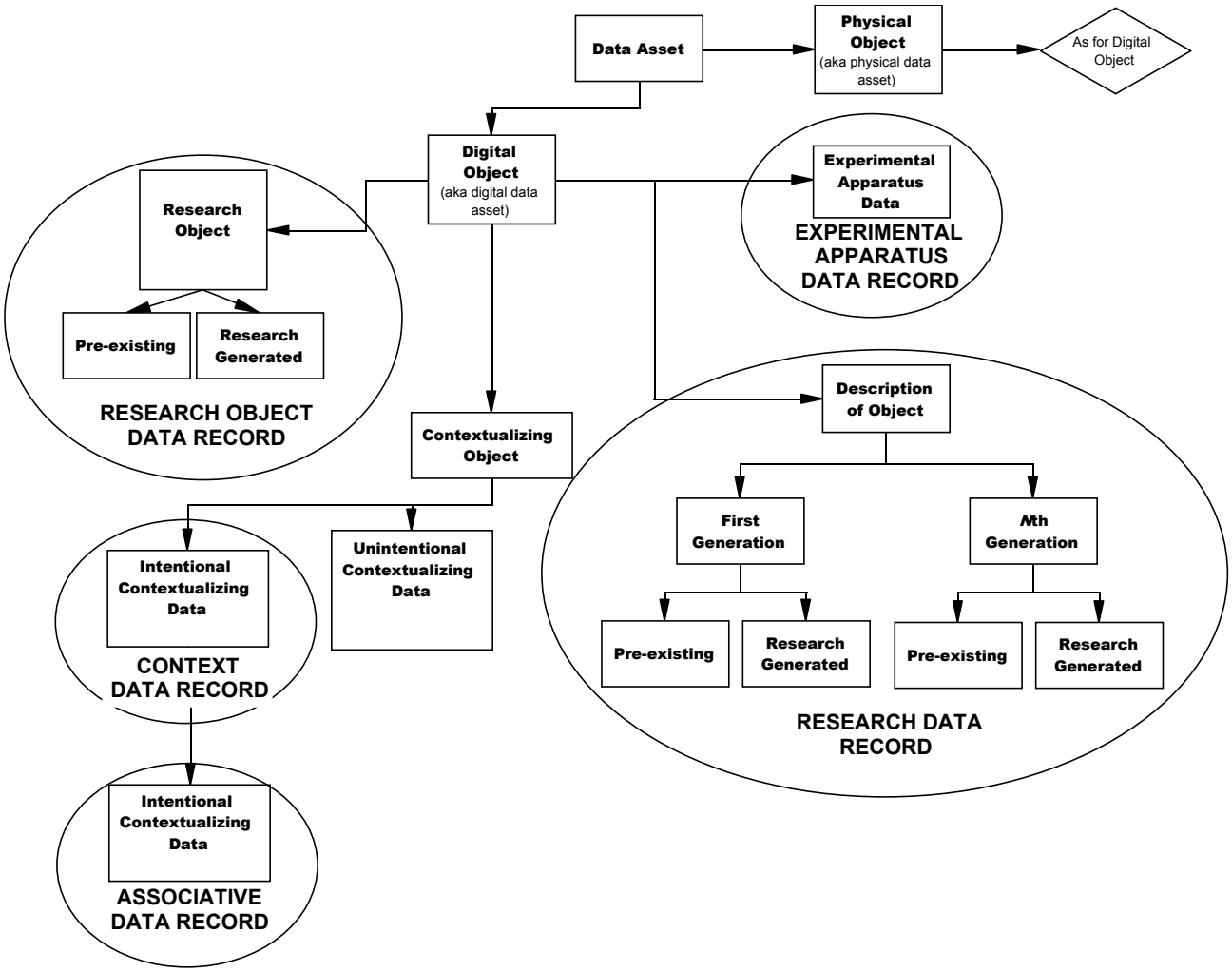

Figure 1. The Research Activity Data Object Taxonomy

The development of the data in a research activity, especially in respect of the data records in which the data are found, can be captured by means of a **RAID Diagram** of the sort shown in Figure 2. This illustration provides, using a ULM (Universal Modelling Language) representation, information about each data record acquired or created during the research activity (in this case a complete research project). The sorts of information that can be provided are such things as record type and format, the rôle played by the record content in the development of data and whether the record was acquired intact as a research artefact or whether it was generated during the course of the research activity. In addition, and most importantly, the relationship between records is shown in terms of the particular data development activity through which they came into being. This mapping provides a contextualizing record for the entirety of data

records created during the research and may be used for a number of purposes.

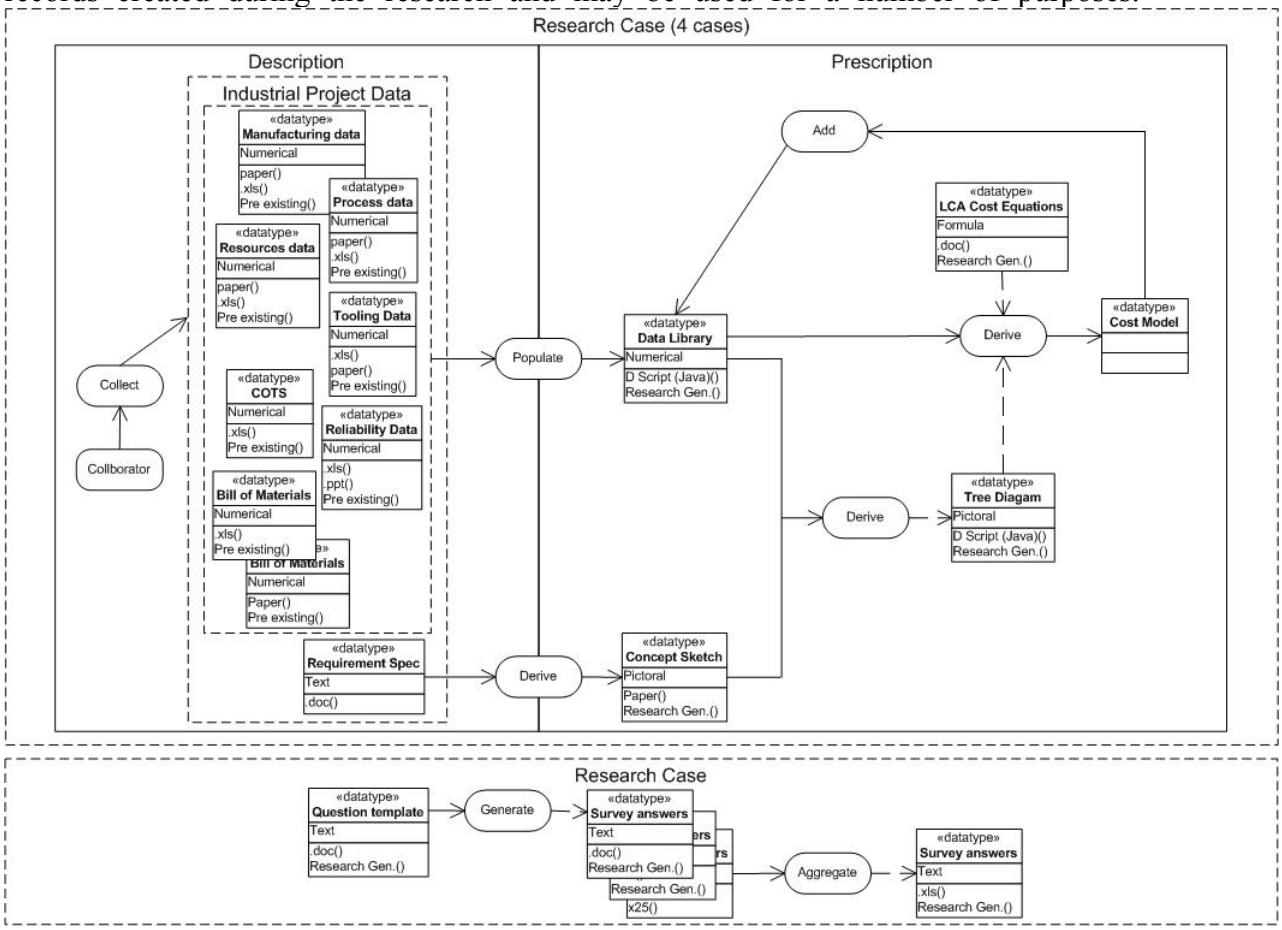

Figure 2. A RAID diagram of a typical research activity, in this case research associated with aerospace cost-modelling

2.1 The benefits of data development mapping

It is clear that the creation and provision of a RAID diagram for a research activity will provide a rich description of the research data assets associated with that research activity. Apart from identifying the existence of useful information about each record as an individual data asset, a diagram may also provide a clear representation of the relationships and ordering between the related items.

Furthermore, the evidence of the existence of and the information contained within those records provided specifically for the purpose of contextualization (that is the *contextual data records* of which the *associative data record* type is represented by the RAID diagram itself) provides a firm basis by which the data can be interpreted and understood. Thus is provided the necessary foundations for re-use and repurposing.

In addition to this, however, are other useful functions that the RAID Diagram for a specific case can provide. Whilst the information contained in the RAID diagram shown above is limited, the RAID modelling approach allows in principle an unlimited amount and diversity of information to be encoded and then harvested about the data development activity. For example, as discussed in Howard, et al, (2010), the act of developing data carries with it the potential for side-effects which may impinge upon information re-purposing and re-use, including importantly information loss and

function loss. By identifying in what way data have been manipulated from one state to another, the potential for side-effects may be predicted, flagged up and measures taken in mitigation. The RAID record might be used also to support the creation of the data records required in the formal processes of data curation and archiving.

3. THE APPLICATION REQUIREMENTS DEFINITION

The fundamental purpose of the proposed application is to support recording of the development of data and Data Records that are generated during a 'research activity'. The archetypal research activity for which the system will be used will be the research project, being a focused activity carried on by one or more individuals; however, any coherent and logically bounded research effort, for example a specific task within a research project, or research activity distinguished by theme, topic or line of enquiry may benefit from data development recording. As such it will be possible to use the system for research activities at different levels of decomposition as is appropriate to the research being carried out. A set of data assets, at whatever level of decomposition is selected, will constitute a Data Case. There may be one or more data cases associated at a less detailed level of decomposition.

The authors believe that ideally information management interventions should result in a zero net resource increase. Ideally, then, the application should be fully automatic, with no intervention necessary from the application user. This ideal is motivated principally in this case by the need to minimize the intrusion into the work flow of things not directly related to the research work that is being carried out and supported. However, it will be necessary – because of the necessary minimal interaction of the user, and the limitations of automatic data gathering and on-the-fly inferencing – to elicit information needed by the system directly from the application user. As a means of minimizing user interaction with the application, as much information as possible will be harvested from the operating system and from metadata contained within, for example, the records themselves. This trade-off will become apparent when the application software design is in progress. It is expected then that a semi-automatic, or user-assisted, system will be implemented.

In any event, the value of the information recorded and the power of the system may be enhanced greatly by allowing the application user to volunteer additional information related to data and data record development, information of the sort that the system would have no access to. This may be seen as wishful thinking, but it is not unforeseeable that a researcher may annotate or add metadata to a data asset in order to make its later interpretation and use easier. In recognition of the above, the system will have two levels of operating 'competence': basic and enhanced. In basic mode *all* functions will be carried out fully automatically without intervention from the application user, except in as much as the user must initiate a particular data-gathering episode and initialize the system with mandatory data. In enhanced mode, the system will elicit information from and will be able to accept and integrate additional non-mandatory information volunteered by the application user. Necessarily, the information gathered in each of these modes will be different as will the richness of the information and thus the captured knowledge about the data. The intention, then, should be to implement an application in which at one and the same time the richness of data recorded is maximized and yet the interaction with the user is minimized.

The system will provide a working environment in which the 'research data development' will be carried out, managed and recorded. It will, thus, constitute both a data management system and an electronic and physical artefact[1] management system. The primary purpose of the system is information capture to support data purposing, repurposing and 'supporting data re-use', but it will also be used to guide the application user in fulfilling data and data record management to the same end.

The system will record the induction into the environment of existing data assets (in the form of data records), and then the instances of data development activity (see Table 1 above) which occur.

The principal reporting mode for the system will be via representation in a diagram of a RAID model of the research activity. Because the representation model is based on UML, the information content can be expanded infinitely. However, the system will minimally provide for the capture and representation of the metadata as described as an annex to the functional specification which follows.

The RAID diagram will be the principal contextualizing representation for the data that is gathered, and will be used as an active interface for interrogation and change or augmentation of the underlying data and, where they are electronic files, for access to the data assets that are represented. It will be possible to select 'data development paths' and home in on information associated only with the development of a selected set of data records. Direct textual access to the underlying data contained within the RAID Record will also be possible.

It is possible, though not intended for current implementation, that the application will be integrated into an environment that will provide other functions for supporting data management, such as annotated metadata capture and the enforcement of such things as file name usage, unique ID allocation and approved-term adoption for data assets. At the same time functional enhancement would be desirable that allowed recording of within-record data development at the data level. This functionality is not, however, contemplated for the current design requirement. Thus, as identified in the use-case analysis, these functions are out-of-scope in respect of the functional requirements of the application.

4. FUNCTIONAL SPECIFICATION

The functional requirements treated in the functional specification below are based upon the use cases extended to take in the first-level system function calls which are logically entailed. Further elaboration of the system functions will be required during the software design phase as too will any extensions to the user requirements. Some further considerations are discussed in Section 0.

[1] It is recognized by the authors (Howard, 2010) and others (e.g. CCS02) that research data may be embodied and thus recorded in physical artefacts. At the same time the concept of 'initiating object' has been introduced as one of the two precursors to data development (see Section 4.1, system function 6). It follows that any system which is able to provide a complete contextualizing record for research data must be able to embrace references to physical objects.

4.1 User-driven Functions

USER FUNCTION 1	**Start a new RAID record**	See Use Case 1
System Function 1	**Launch application interface**	Basic interface framework
System Function 2	**Check user credentials**	Data development records need to be secure against uninformed tinkering
System Function 3	**Display initiating user action**	1) Open RAID Record, or 2) New Record
System Function 4	**Open existing RAID record**	Open file view
System Function 5	**Start new RAID record**	System response is to populate the record mandatory metadata from system-based information, based on user credentials then prompt user for missing mandatory information.
System Function 6	**Open RAID Record metadata dialogue box**	Ordered data: prompt for **Initiating Object** or **Initiating Event** and complementary data record. System accepts missing data
System Function 7	**Prompt user to save**	Standard system function initiated by user action or on failure-to-save-before-closing
System Function 8	**Save to RAID record**	Executed on user demand
System Function 9 (see User Function 3)	**Get RAID diagram**	Executed as system demand on completion of save

USER FUNCTION 2	**Get RAID Record Metadata Digest**	See Use Case 2
System Function 10	**Display metadata options dialogue box**	Option 1) RAID record; option 2) Data Record
System Function 11	**Display RAID record file list**	System accepts user navigation and selection of one file
System Function 12	**Print to screen metadata digest table**	

USER FUNCTION 3	**Get a RAID diagram**	See Use Case 3
	Get a RAID diagram	See System Function 3
System Function 11	**Display RAID record file list**	System accepts user navigation and selection of one file
System Function 14	**Print to Screen RAID diagram**	

USER FUNCTION 4	**Check mandatory metadata Set**	See Use Case 4
	Display metadata options dialogue box	See System Function 10
System Function 15	**Display data record file list**	System accepts user choice and selection of one file
System Function 16	**Display all-metadata dialogue box**	
System sub-Function 17	**Display all-metadata dialogue box with mandatory fields highlighted**	
System Function 18	**Validate entry data type**	Sub-routines include error checking and rejection/acceptance
	Prompt user to save	See System Function 28
	Save to RAID record	See System Function 8

USER FUNCTION 5	**Check Completeness/status of a data record**	See Use Case 5
	Display metadata options dialogue box	See System Function 10; user selects ‘data record’
	Display data record file list	See System Function 15
System Function 13	**Display two-choice menu {mandatory metadata set; all metadata}**	User selects choice
	Display metadata dialogue box	See System Function 16 or System Function 17
	Validate entry data type	See System Function 18
	Prompt user to save	See System Function 19
	Save to data record	See System Function 8

USER FUNCTION 6	**Annotate data record through a RAID Diagram**	See Use Case 6
System Function 19	**Display element data dialogue box**	
	Validate entry data type	See System Function 18
	Prompt user to save	See System Function 28
	Save to RAID record	See System Function no 8

USER FUNCTION 7	**Add Data Record to RAID Record**	See Use Case 7
System Function 20	**Precursor dialogue box**	Allows selection of precursor to the data record: 1) Existing object 2) initiating object or 3) initiating event
System Function 21	**Display list of precursor objects**	Precursor objects are any of those things from which a data record can be developed
System Function 22	**Add precursor object**	Precursor objects: 1) pre-existing data record, 2) pre-existing physical object
System Function 23	**Add precursor event**	
System Function 24	**Display data record all-metadata dialogue box: mandatory fields highlighted**	This provides a dialogue box for the metadata required for a single data record
	Validate entry data type	See System Function 18
	Prompt user to save	See System Function 28
	Save to RAID record	See System Function 8

USER FUNCTION 8	**Update Record Metadata**	See Use Case 8
	Display metadata options dialogue box	See System Function 10; user selects ‘data record’
	Display data record file list	See System Function 15
	Display all-metadata dialogue box	See System Function 17
	Validate entry data type	See System Function 18
	Prompt user to save	See System Function 28
	Save to RAID record	See System Function 8

USER FUNCTION 9*	**Associate a data record with another data record**	See Use Case 9
System Function 25	**Display data record file list**	See System Function 15 User can select one file from list
System Function 26	**Set association type**	The association type is taken from a list of authorized relations
	Display data record file list	See System Function 15; user can select one file from list
	Prompt user to save	See System Function 28
	Save to RAID record	See System Function 8
System Function 27	**Save metadata to data record**	This function acts on the previously selected data records

*** Note:** This activity could also be carried out using the 'annotate RAID diagram' function as a sub-routine.

USER FUNCTION 10	**Associate one Data Case with another**	See Use Case 10
	Display RAID record file list	See System Function 11
System Function 28	**Prompt user to select *n* items**	$n > 2$
System Function 29	**Name-and-save file**	
	Display RAID diagram	See System Function 14

USER FUNCTION 11	**Fragment RAID record**	**See Use Case 11**
Graphical Function 1	**Select annotate diagram function**	
Graphical Function 2	**Select and highlight graphical items**	
	Name-and-save file	See System Function 29
System Function 30	**Prompt user to delete highlighted graphical functions**	This function will zap all highlighted graphical elements and update the view.

USER FUNCTION 12	**Select a Sub-set of Data Records in a RAID Diagram for Inspection**	See Use Case 12
	Select and highlight graphical items	See Graphical Function 1
	Display metadata digest	Metadata displayed for selected elements

USER FUNCTION 13	**Associate and label a sub-set of data records through a RAID Diagram**	See Use Case 13
	Select annotate diagram tool	See Graphical Function 1
Graphical Function 3	**Select boundary tool and place box round data record entities**	
Graphical Function 4	**Display boundary box dialogue box**	
	Validate entry data type	See System Function 18
	Prompt user to save	See System Function 28
	Save to RAID record	See System Function no 8

4.2 Additional considerations relating to the implementation

There are many ways in which the functionality could be extended given the resources. The following topics are presented for consideration:

1. As identified earlier the ideal for the application is to tend toward full automation. Much of the functionality could be carried out and enhanced based on a systems watch of the elements in a RAID record, and the metadata available to the system based on system activity and metadata contained in each of the data records mapped. This approach would be particularly necessary were the intra-record development of data to be recorded at the data level. Such functions as datum import and data manipulation tracing would allow a detailed record to be developed of how the data in record has changed and provide a basis for flagging up management side-effects relating to such things as data loss, information loss and function loss.
2. The inferencing power of the system could be enhanced substantially if it were to be underpinned by a research data ontology. Given an ontology and suitable axioms, much of the associative information could be acquired on the fly.
3. The inferencing power of the system could be enhanced substantially if it were to be embedded in a larger research data management support environment and had access to the systems data gathered and generated by that environment.
4. As a RAID record develops (particularly if system-led) the opportunity will arise for the system to flag management issues that arise and to prompt the user to take actions which will enhance the usefulness of the record. An example of this is in reminding the user of the side-effects of data development and suggesting ameliorating actions.

5. Where a large repository of research data exists, a search facility pointed at the RAID Record metadata would enhance research data rediscovery for re-use.

4.3 Considerations relating to metadata requirements and exchange

The RAID record for a data case will contain a large and diverse information content which describes the RAID record itself and the data assets (that is the diversity of data records within the data case). There are two principal sets of metadata important to the operation of the associative tool. There are those which are tool-specific, derived from the functional specification, and those which are record-specific and which describe the record as individual resources or objects.

In the software design phase consideration should be given to specifying the minimum data assets that are considered necessary for the useful functioning of the application in respect of the different types of data object being considered. This minimum requirement for each type of data object is referred to in this requirement specification as the mandatory metadata set. For flexibility it is suggested that the mandatory metadata set should be bespoke; that is to say it can be specified, by for example a data manager, to suit the research context in which the tool is being used and the prevailing data management protocols.

A question arises also as to what extent metadata useful to the RAIDmap application should be shared between the application (within a RAID record) and the data records themselves, and whether a specification required within-data-record metadata should become part of the application requirement in order for the application functions to be properly supported. Data records which have a rich metadata content would assist in the quest for greater automation of the system. The question then arises as to whether the within-data-record metadata should be compiled into the RAID record or merely be available on demand to the system when a RAID record (and its concomitant data records) is in scope. These matters will require consideration during software design.

The suggested set of implementation-independent metadata is given in Appendix 1. The set is given only in respect of the RAIDmap application requirements and is not considered to identify all metadata elements that would be required for a complete implementation of a associative tool, but that suggested by the functional requirements. In particular the additional metadata requirement of the record (aka resource) will need to be taken into account in respect of those aspects of curation and archiving that are not embraced by the requirements for *re-purposing* and *supporting data re-use* as defined here. The selection of the general metadata has been influenced by and draws on existing resources such as the Dublin Core Metadata Element Set (DCMI), the Scholarly Works Application Profile (SWAP) and PREMIS (see Ball (2010) for identification and a discussion of such metadata resources)), the Metadata Open Description Schema (MODS) (Gartner, 2003) and DataCite (Starr & Gastl, 2011). In addition to this it is likely that consideration will need to be given to adopting a metadata schema which will provide for the provision of recording information relating to rights, licensing, contractual and confidentiality issues (as supported by, for example, the Open Digital Rights Language (2008)).

4.4 Non-functional requirements

This requirement has not considered non-functional requirements, such as supporting platform and portability, cost or user-base size, etc. The single non-functional requirement recorded here is that the RAIDmap application file type extension should be .rmap.

5. GLOSSARY

Note: The following terms are taken from the more complete ERIM Terminology of terms associated with data re-use and re-purposing which can be found here.

Associate. To make explicit the relationship between items of Data, Data Records or Data Cases.

Associative Data Record A Context Data Record containing Associative Data which records the association between other data records or data.

Context Data Record (CDR). A Record containing Context Data.

Research Object Data Record. A data object (being either a physical or electronic data record) containing data which is, itself, the object of the research enquiry.

Experimental Apparatus Data Record. A data object (being either a physical or electronic data record) containing symbolic representations which are functionally analogous to the physical experimental apparatus familiar in much laboratory-based research.

Context Data. Data that support the Research Activity but do not describe the research object nor are the research object itself.

Data Case The set of Data Records (Research Data Records and Context Data Records) associated with some discrete Research Activity (project, task, experiment, etc.).

Data Object. Either a Physical Object or a Digital Object.

Data Record (DR). The Data Object which contains the Data.

Data. Reinterpretable representations of information in a formalized manner suitable for communication, interpretation or processing.

Initiating Event Any happening subject to research scrutiny and which is the precursor to the recording of data as manifest in a data record. Examples of events include interviews, meetings design episodes and so on.

Initiating Object Any object subject to research scrutiny and which is the precursor to the recording of data as manifest in a data record. Such objects include physical objects from which data are generated or from which data is obtained by observation or, singularly, a Research Object Data Record.

RAID Diagram A 2-D graphical representation based on the RAID Modelling method of the data records within a data case and their associations and principal characteristics. The RAID record visualizes a part of the data contained within the RAID Record.

RAID Modelling A method of modelling the development of data during the research activity which identifies data records, their relations, associations and metadata.

RAID Record The underlying record of a data case containing information about data records constituting the case, their temporal ordering, relations and association and principal characteristics. Also a data file in which the record is contained identified as such by the RAIDmap application suffix .rmap.

Research Activity. The process through which Research Data and Context Data are accumulated and developed.

6. REFERENCES

Alex Ball. (2010). *Review of the State of the Art of the Digital Curation of Research Data* (version 1.0). ERIM Project Document erim1rep091103ab10. Bath, UK: University of Bath.

CCS02. Consultative Committee for Space Data Systems (2002). *Reference Model for an Open Archival Information System (OAIS)*. Blue Book CCSDS 650.0-B-1. Also published as ISO 14721:2003. URL: http://public.ccsds.org/publications/archive/650x0b1.pdf.

Cobos, R. and Schlichter, J. (2004). Collaborative Knowledge Transfer by Annotating Documents, in D. Karagiannis and U. Reimer (Eds.): *Practical Applications of Knowledge Management*, 2004, Lecture Notes in Computer Science, Vol3336/2004, 153-158, DOI: 10.1007/978-3-540-30545-3_14.

Fancellu, D. (n.d.). Functional Specification Standard, URL: http://www.softwarereality.com/lifecycle/functionalspec.jsp. Downloaded 15 October 2010.

Howard, T., Darlington, M., Ball, A., McMahon, C. and Culley, S. (2010). *Understanding and Characterizing Engineering Research Data for its Better Management* (version 1.0). ERIM Project Document erim2rep100420mjd10. Bath, UK: University of Bath.

Gartner, R, (2003). *MODS: Metadata Object Description Schema,* JISC Technology and Standards Watch Reports 2003. URL: http://www.jisc.ac.uk/uploaded_documents/tsw_03-06.pdf.

Kelly, B. (2004) *Managing And Using Metadata In An E-Journal*,., QA Focus case study 01, UKOLN, http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-01. Downloaded 01 February 2011.

Open Digital Rights Language. (2008) version 1.1. URL: http://odrl.net/1.1/ODRL-11.pdf. Downloaded 26 January 2011.

Pressman, R. (2005). *Software Engineering: A Practitioner's Approach*, 6th Ed. R.S. Pressman and Associates, ISBN: 0072853182, p. 61.

Sefton, P., Barnes, I., Ward, R. & Downing, J. (2009). Embedding Metadata and Other Semantics In Word Processing Documents. *International Journal of Digital Curation*, Vol 4, No 2, pp93-106.

Starr, J. and Gastl, A. (2011), 'isCitedBy: A Metadata Scheme for DataCite', *D-Lib Magazine* **17** (1/2) . DOI:10.1045/january2011-starr.

APPENDIX 1 ASSOCIATIVE TOOL SUGGESTED METADATA ELEMENTS

The metadata elements listed here are those suggested as being at least useful in relation to the Research Data Associative Tool (RAIDmap). They have been derived from the associative tool functional specification and drawn from a number of sources as indicated in the table against the appropriate metadata entries. A complete list of elements would be augmented and confirmed during the software design process.

Attribute Label	RAID-related Definition	Value set	Type	Comments
RAID Functional Specification				
Data development process type	One of a set of processes that are commonly carried out during the Research Activity which changes or adds to the Research Data associated with a research activity or project.	addition, association, aggregation, annotation, augmentation, collection, collation, deletion, derivation, duplication, extraction, generation, migration, population, refinement	Expanding set	These relations could be encoded using terms similar to those defined in IsCitedBy, e.g. isAggregationOf, isCollationOf, the value being the data record Identifier
Data record type	One of five types of data record specified in the RAID model	research, research object, experimental apparatus, context, associative,	Fixed set	
Data object type	Either a Physical Object or a Digital Object.	digital, physical	Fixed set	
RAID datatype description	Content-level classification of a data record	requirement specification, bill of materials, manufacturing data, data records data, cost data, reliability data, cost model, concept sketch, data library, question template, survey response, data record data, tooling data,	Descriptive expanding set	
1st-Generation Data Type	The Data resulting either from Data Collection or from Data Generation.	pre-existing, research-generated	Fixed, binary	
Media type		text, numerical, pictorial_2D,	Expanding set	

Attribute Label	RAID-related Definition	Value set	Type	Comments
		pictorial_3D, video, audio,		
Data format	The organization of digital information according to a preset (sometimes application-specific) specification	.doc, etc	Expanding set	
Reality	Event classification reflecting conditions in which research data were collected	real, simulation	Fixed, binary	
Stability	Dataset status relating to expectation of future dataset development	open {expanding, dynamic, expanding&dynamic}; closed {definitive, dormant}	Binary, sub-qualifiers	
Collection method	The means by which the research data were collected, generated or recorded	questionnaires, field note, discussion transcription, experimental rig output, commercial software, bespoke commercial software, fully bespoke software, discursive process recording {brainstorm, discussion, brain dump, …}	Expanding set	
Repeatability	A measure of the practical possibility of a Generation process being repeated such that Data Reproducibility is possible in principle	repeatable, non-repeatable	Binary	
Interpretation	Indication of whether derived data were the result of objective or subjective interpretation of research data	Objective, subjective	Binary	

Note that some of the attributes consist of necessarily incomplete or expanding sets, and some require bespoke descriptive content (which would require a free text field).

Data record (aka data object)

Attribute Label	RAID-related Description	Value set	Type	Comments
Dublin Core Sub-set				
Creator	Probably the researcher or research team responsible for making or collecting the data record	N/a	Free text	
Description	A brief account	N/a	Free text	
Date	The date of acquisition/generation into the data case	N/a	Free text	
Format, see Data Format above		N/a	Free text	
Identifier	An unambiguous reference to the data record within a given context; preferably based on a formal naming system such as a file naming protocol	N/a	Free text	Consideration should be given to the context agreed for the 'unambiguous reference'. It is unlikely that globally-unique identifiers would be appropriate; rather that a project-based context would suffice.
Source	A Reference to a data record from which the present data record is derived through a data development process	–	DC:Identifier	
Subject	The topic represented in keywords		Free text	
Title	A name given to the data record		Free text	
PREMIS sub-set				
Creating application	The application which created the	n/a	Free text	

Attribute Label	RAID-related Description	Value set	Type	Comments
	data record, together with version			
originalName	The name of the pre-existing object as submitted to or collected by the researcher (team), before any renaming has been carried out	n/a	Free text	This applies only to pre-existing data records
storageMedium	The physical medium on which the object is stored	Magnetic tape, hard disk, CD-ROM, DVD, flash memory. etc	Expanding list	This information could make it easier to locate the data asset
Dependency	Information about a non-software component or associated file needed in order to use or render the representation or file, for example, a schema, a DTD, or an entity file declaration.	n/a	Free text	
Relationship	Information about a relationship between one data record and one or more other data records	–	relationshipSubType	From, PREMIS **1.10 relationship Semantic components** 1.10.1 relationshipType 1.10.2 relationshipSubType 1.10.3 relatedObjectIdentification 1.10.4 relatedEventIdentification
relatedEventIdentification				
DataCite v2.0 sub-set				
Relationship sub-type:		n/a		These may be used simultaneously in relation to a single

Attribute Label	RAID-related Description	Value set	Type	Comments
				data record
IsSupplementTo	Use to indicate the relation to the data record to which this data record is a supplement.	–	DC:identifier	
IsSupplementedBy	Use to indicate the relation to the data record(s) which are supplements of this data record.	–	DC:identifier	
IsContinuedBy	Use to indicate the data record is continued by the data record referenced by the related identifier.	–	DC:identifier	
Continues	Use to indicate the data record is a continuation of the data record referenced by the related identifier.	–	DC:identifier	
IsNewVersionOf	Use to indicate the data record is a new edition of an old data record, where the new edition has been modified or updated.	–	DC:identifier	
IsPreviousVersionOf	Use to indicate the data record is a previous edition of a newer data record.	–	DC:identifier	
IsPartOf	Use to indicate the data record is a portion of another data record.	–	DC:identifier	
HasPart	Use to indicate the data record is a container of another data record.	–	DC:identifier	
IsReferencedBy	Use to indicate the data record is used as a source of information by another data record.	–	DC:identifier	
References	Use to indicate the relation to the data record which is used as a source of information of the data record.	–	DC:identifier	

Attribute Label	RAID-related Description	Value set	Type	Comments
IsDocumentedBy	Use to indicate the data record is documentation about/explaining the data record referenced by the related identifier.	–	DC:identifier	Limited to context data records or associative data records
Documents	Use to indicate the relation to the data record which is documentation.	–	DC:identifier	
IsCompiledBy	Use to indicate the data record or data is compiled/created by using another data record or dataset.	–	DC:identifier	
Compiles	Use to indicate the data record is used for creating another data record or dataset.	–	DC:identifier	
isVariantFormOf	Use to indicate the data record is a variant or different form of another data record, e.g. calculated or calibrated form or different packaging.	–	DC:identifier	
isOriginalFormOf	Use to indicate the relation to the data records which are variant or different forms of this data record.	–	DC:identifier	